

IN THE CLAIMS:

Please amend the claims as set forth in the following listing. This listing of claims will replace all prior versions, and listings, of claims for the present application:

1. (Currently amended) A system for allocating resources amongst a plurality of applications running in a distributed, multiprocessor computing environment, the system comprising:

a plurality of server computers in a server pool, wherein each server comprises;

a distributed server pool director for organizing and maintaining [[a set]] the plurality of servers in [[a]] the server pool;

a monitoring module at each server computer in the server pool for detecting demands for one or more resources located on the server computer and exchanging information amongst the plurality of server computers regarding demands for the one or more resources at the plurality of servers;

a distributed policy engine for specifying a policy for allocation of resources of the plurality of server computers amongst the plurality of applications having access to the resources, wherein the distributed policy engine bases the policy is based on the demands for the one or more resources, monitors the system to determine when one or more conditions exist, and adapted adapts the policy based on changes in the behavior of the one or more conditions, the system and wherein the policy engine is distributed on each server computer in the plurality of server computers; and

an enforcement module at each server computer for allocating the resources amongst the plurality of applications based on a decision made by the policy engine.

2. (Currently amended) The system of claim 1, wherein the resources include one or more of communication resources, network bandwidth, processing resources, memory, disk space, system I/O (input/output), printers, tape drivers, and software licenses,
wherein one or more conditions comprises the existence of another program and interdependencies between applications running on the system.

3-5. (Canceled).

6. (Currently amended) The system of claim 1, wherein the policy engine receives user input for defining an application subject to the policy, wherein the user input includes defining components of an application.
7. (Currently amended) The system of claim 6, wherein the monitoring module identifies an application running at a given server computer based, at least in part, upon the user input for defining the application.
8. (Currently amended) The system of claim 7, wherein the monitoring module detects a request for resources by the application at the given server computer.
9. (Canceled).
10. (Currently amended) The system of claim [[9]] 6, wherein the components include a selected one of processes, network traffic, and J2EE components.
11. (Currently amended) The system of claim 1, wherein the policy engine receives user input of a policy specifying actions to be taken for allocation of the resources in response to particular conditions, wherein the policy includes one or more of an attribute indicating when a particular condition of the policy is to be evaluated, a command to be run in response to a particular condition, an attribute indicating when action is to be taken based upon a particular condition of the policy being satisfied and priorities of the plurality of applications to the resources.
- 12-15. (Canceled)
16. (Currently amended) The system of claim [[15]] 11, wherein the enforcement module allocates resources amongst the plurality of applications based, at least in part, upon the specified priorities.
17. (Currently amended) The system of claim 1, wherein the policy engine includes a user interface for a user to specify the policy one or more priorities of the plurality of applications.

18. (Original) The system of claim 1, wherein the policy engine supports an expression language for policy definition.

19. (Canceled).

20. (Currently amended) The system of claim 1, wherein the monitoring module performs one or more of determines determining resources available at each server computer, determining resource utilization at each server computer and exchanging resource utilization information amongst the plurality of computers.

21-22. (Canceled).

23. (Currently amended) The system of claim 1, wherein the enforcement module performs one or more of allocates allocating network bandwidth amongst said plurality of applications based upon the policy and information regarding demands for the resources, allocating processor resources amongst said plurality of applications based upon the policy and information regarding demands for the resources and communicating with an external module for specifying allocation of resources by said external module.

24-25. (Canceled).

26. (Currently amended) The system of claim ~~[[25]]~~ 23, wherein said external module includes one or more of a load balancer for load balancing instances of an application, a router and a provisioning device.

27. (Canceled).

28. (Currently amended) The system of claim 1, wherein the enforcement module starts an instance of an application on a given server computer based upon the policy and information regarding demands for the resources.

29. (Withdrawn) An improved method for allocating resources of a plurality of computers to a plurality of applications, the method comprising:

receiving user input specifying a dynamically configurable policy for allocating resources of a plurality of computers amongst a plurality of applications having access to the resources;

at each of the plurality of computers, detecting demands for the resources from the plurality of applications and availability of the resources;

exchanging information regarding demand for the resources and availability of the resources amongst the plurality of computers; and

allocating the resources to each of the plurality of applications based on the dynamically configurable policy and the information regarding demand for the resources and availability of the resources.

30. (Withdrawn) The method of claim 29, wherein the resources include communication resources.

31. (Withdrawn) The method of claim 30, wherein the communication resources include network bandwidth.

32. (Withdrawn) The method of claim 29, wherein the resources include processing resources.

33. (Withdrawn) The method of claim 29, wherein the resources include a selected one of memory, disk space, system I/O (input/output), printers, tape drivers, load balancers, and software licenses.

34. ((Withdrawn) The method of claim 29, wherein said receiving step includes receiving user input for defining an application.

35. (Withdrawn) The method of claim 34, further comprising:

providing a set of default rules for assisting a user in defining an application.

36. (Withdrawn) The method of claim 29, wherein said detecting step includes detecting applications running on each of said plurality of computers.

37. (Withdrawn) The method of claim 29, wherein said detecting step associating particular processes with a particular application.

38. (Withdrawn) The method of claim 29, wherein said detecting step associating particular network traffic with a particular application.
39. (Withdrawn) The method of claim 29, wherein said detecting step includes detecting components of an application.
40. (Withdrawn) The method of claim 39, wherein said step of detecting components of an application includes detecting a J2EE component.
41. (Withdrawn) The method of claim 29, wherein said receiving step includes receiving user input specifying actions to be taken for allocation of the resources in response to particular conditions.
42. (Withdrawn) The method of claim 41, wherein the user specifies a script to be run based upon a particular condition.
43. (Withdrawn) The method of claim 41, wherein the user specifies when a particular condition of the dynamically configurable policy is to be evaluated.
44. (Withdrawn) The method of claim 43, wherein the user specifies when action is to be taken based upon the particular condition being satisfied.
45. (Withdrawn) The method of claim 29, wherein said receiving step includes receiving user input specifying priorities of the plurality of applications to the resources.
46. (Withdrawn) The method of claim 45, wherein the allocating step includes allocating resources amongst the plurality of applications based, at least in part, upon the specified priorities.
47. (Withdrawn) The method of claim 29, wherein the receiving step includes providing an expression language for policy definition.
48. (Withdrawn) The method of claim 29, wherein said detecting step

includes determining resource utilization at the given computer.

49. (Withdrawn) The method of claim 48, wherein said determining resource utilization step includes determining average resource utilization over a given time period.

50. (Withdrawn) The method of claim 29, wherein said allocating step includes allocating a specified amount of resources to a particular application when the particular application is initially detected at a given computer.

51. (Withdrawn) The method of claim 29, wherein said allocating step includes allocating resources based upon particular events.

52. (Withdrawn) The method of claim 29, wherein said allocating step includes allocating network bandwidth to each of the applications.

53. (Withdrawn) The method of claim 29, wherein said allocating step includes allocating processor resources amongst the plurality of applications.

54. (Withdrawn) The method of claim 29, wherein said allocating step includes communicating with an external module for allocating resources provided by an external module.

55. (Withdrawn) The method of claim 54, wherein the external module comprises a load balancer for load balancing in stances of an application.

56. (Withdrawn) The method of claim 54, wherein the external module comprises a selected one of a router and a provisioning device.

57. (Withdrawn) The method of claim 29, wherein said allocating step includes starting an instance of an application on a given computer.

58. (Withdrawn) A computer-readable medium having processor executable instructions for performing the method of claim 29.

59. (Withdrawn) A downloadable set of processor-executable instructions for performing the method of claim 29.

60. (Currently amended) A method for allocating resources to a plurality of applications, the method comprising: providing a plurality of communicatively connected servers in a server pool; at a policy engine distributed amongst the plurality of communicatively connected servers in a multiprocessor computing environment,

receiving user input specifying priorities of the plurality of applications to resources of [[a]] the plurality of servers, the specified priorities including designated servers assigned to at least some of the plurality of applications;

selecting a given application based upon the specified priorities of the plurality of applications;

determining the demand for one or more resources located on each server in [[a]] the plurality of communicatively connected servers;

specifying a policy for allocation of resources of the plurality of servers;

allocating to the application resources located on one or more of the plurality of servers based on the policy for allocation of resources, wherein a server pool director organizes and maintains the plurality of servers;

allocating additional resources to the application until the application's demands for resources are satisfied, wherein a local workload manager component on each server regulates the resources on that server based on the allocation of resources determined by the policy engine; and

repeating above the steps of determining the demand for one or more resources, specifying a policy for allocation of resources, and allocating resources to the application for each of the plurality of applications based on the specified priorities.

61. (Original) The method claim 60, wherein the receiving step includes receiving user input of a value for a given application representing relative priority of the given application compared to other applications.

62. (Canceled).

63. (Currently amended) The method of claim [[62]] 60, wherein the step of selecting a given application includes commencing with selection of an application having the highest priority.

64. (Currently amended) The method of claim 60, further comprising one or more steps of:

powering on a server allocated to an application if the server is in a powered off state,

adding a server newly allocated to an application to a set of servers across which the application is allocated,

removing a server no longer allocated to an application from a set of servers across which the application is allocated,

determining whether an application is inactive on a server allocated to the application and initiating a resume script for running the application on the server if the application is determined to be inactive,

determining whether a server no longer allocated to an application is in a set of servers designated for the application and running a suspend script if the server is determined to be in the set of servers, and

if a suspend script is executed on the server, determining whether the server should be powered off based on consulting a power management rule and powering off the server if it determined that the server should be powered off.

65-69. (Canceled)

70. (Original) The method of claim 60, wherein said allocating step includes starting an instance of an application on a given computer.

71. (Currently amended) A system comprising: a server pool director for organizing and maintaining a plurality of servers, wherein the server pool director is distributed across a plurality of communicatively coupled servers, wherein each server comprises a processor and a computer-readable medium having processor executable instructions for performing the method of claim 60, that, when executed by the processor, perform the steps of:

by a policy engine distributed across the plurality of servers,

receiving user input specifying priorities of the plurality of applications to resources of the plurality of servers, the specified priorities including designated servers assigned to at least some of the plurality of applications;

selecting a given application based upon the specified priorities of the plurality of applications;
determining the demand for one or more resources located on each server in the plurality of communicatively connected servers;
specifying a policy for allocation of resources of the plurality of servers;
allocating to the application resources located on one or more of the plurality of servers based on the policy for allocation of resources;
allocating additional resources to the application until the application's demands for resources are satisfied, wherein a local workload manager component on each server regulates the resources on that server based on the allocation of resources determined by the policy engine; and
repeating the steps of determining the demand for one or more resources, specifying a policy for allocation of resources, and allocating resources to the application for each of the plurality of applications based on the specified priorities.

72. (Canceled).